DOCUMENT RESUME

ED 279 998 CS 008 691

AUTHOR Willson, Victor L.
TITLE Methodological Limitations of the Application of Expert Systems Methodology in Reading.
PUB DATE Dec 86
NOTE 21p.; Paper presented at the Annual Meeting of the National Reading Conference (36th, Austin, TX, December 2-6, 1986).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Evaluation Problems; *Reading Research; *Research Methodology; *Research Problems; *Research Utilization; *Theory Practice Relationship

ABSTRACT

Methodological deficiencies inherent in expert-novice reading research make it impossible to draw inferences about curriculum change. First, comparisons of intact groups are often used as a basis for making causal inferences about how observed characteristics affect behaviors. While comparing different groups is not by itself a useless activity, progressing directly to training is premature at best. Second, the think-aloud protocol technique is often used for inferring a subject's cognitive structure of subject matter. This method is inappropriate because it assumes that the organization of this structure resides consciously in a person's mind and can be verbally reproduced. Third, retrospective methods have been employed to infer causality by selecting groups currently differing and discovering differences in their past on putative causal variables, which are then inferred to have caused the present differences. While this technique must be used in historical analyses, it becomes suspect when the inferences are used to speculate on implications for current practice. Finally, techniques employed in naturalistic inquiry often confuse a change in methodology with a change in the discipline being studied, and rely heavily on impressionistic, one-shot observation for many facts. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED279998

Methodological limitations of the application of expert systems methodology in reading

Victor L. Willson

Texas A & M University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Victor L. Willson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the Annual Meeting of the National Reading Conference, Austin TX, December 1986

CS008691

Methodological limitations of the application
of expert systems methodology in reading

The comparison between the expert and inexpert, whether in organizations or in humans, has become a major research technique in education in the last decade. In research on problem solving comparisons have been made between physicists and physics students. In reading research the basic comparison is between good and poor readers for much of the information-processing theory being advanced. In educational administration successful schools are compared with unsuccessful schools. Creative students are compared with normal students. In all of these areas a similar paradigm is being used: the expert system is compared on a number of attributes with a less or inexpert system. Some attributes are assumed to be causal (such as program differences or strategies employed) and others are assumed to be outcomes, such as achievement or time to solution of a problem. Differences between the expert and inexpert on the causal variables are then assumed to be evidence for causation of the variables, and these salient variables are promoted as efficacious for remedying the deficiencies of the inexpert. This paradigm is an important departure from the dominant experimental paradigm used in education, and this paper presents a critical examination of its methodological limitations.

It is an interesting situation that the expertise methodology has sprung from two quite different areas of research, cognitive psychology and curriculum behaviorism. Cognitive psychology has drawn from and has itself influenced artifi-

cial intelligence (AI) research in machine computing. As Chi, Feltovitch & Glaser (1982) has noted a shift took place in AI research from power strategies to knowledge strategies. The best knowledge strategies available for study were found in human beings, so that new computer programs were developed to imitate the way human experts organized and processed information, as could be determined in comparisons with inexpert humans. The effective schools movement, not directly influenced by artificial intelligence in any obvious way, sought to study the school environment. One outcome of such study, conducted by anthropologists, psychologists, and educational researchers, was that some schools (classrooms, teachers, administrators, etc.) were superior in performance to others. Comparisons between variables defined as input, or causal, and output or dependent, led to prescriptions for change in inexpert schools to make them more like the best schools being observed. For example Clark & McCarthy (1983) reported on a cohort sequential type design in which volunteer New York City schools implemented a new program based on effective schools literature.

Since most of the expert-novice comparisons were presented in the methodological trappings of experimental research (ANOVA statistical analysis and interpretation) there have been only one or two serious evaluations of the causal logic underlying the method and its basis for making causal inferences (Rowan, Bossert & Dwyer,1983). The main thesis presented here is that all research applying this method suffers from internal validity flaws sufficiently serious to render it uninterpretable. Furthermore,

expertise method is incapable of supporting causal inference regarding change in the inexpert without true experimental research.

Techniques employed

The techniques being drawn upon in research on expertise include, but are not limited to, the following: comparisons between inact groups; think-aloud protocol; naturalistic inquiry, including ethnographic field method; and retrospective research. There are researchers who are employing experimental research as part of their research strategies, and their applications, specifically exempted from the criticism levelled here, will be mentioned as exemplars of appropriate or adequate research.

Comparison between intact groups. This technique is widely used in expertise research. In problem-solving research Chase & Simon (1973) compared the ability of chess masters and novices to chunk board groupings. They found more elements in the chunks of masters than in the novices. Simon & Simon (1978) found differences in the problem-solving behaviors of physicists and physics students using verbal protocols of the tasks each performed when solving novel problems. The effective schools movement has used comparisons between schools defined as outstanding or excellent and those defined as inferior or deficient to make programmatic decisions about how schools ought to be run. A well-constructed criticism of the effective schools research methodology was made by Rowan, Bossert, & Dwyer (1983). Their specific points will be incorporated into this review; these points include difficulty

with causal ordering, instrumentation, limitations of generalization, and nonequivalent control group comparisons.

The effective schools research of the 1970's was employed in examining reading at both the school and classroom level. Teacher effectiveness has been particularly emphasized (Rupley, Wise, & Logan, 1986). Brophy's (1973) work on process-product research with primary grade teachers is a widely cited example; later important studies include Medley (1977) and Rosenshine (1978); the latter study raised the problem of little experimental verification for effectiveness research. The Stanford Program on Teaching Effectiveness (Crawford, Gage, Corno, Staybrook, Mitman, Schunk, Stallings, Baskin, Hanvey, Austin & Newman, 1978), and the First Grade Reading Group Study (Anderson, Evertson & Brophy, 1978) are experimental or quasi-experimental studies based on initial observations contrasting good and poor teachers (Rupley et al, 1986). It is important to note that in these studies curriculum recommendations were made after comparative intervention was made, not directly on the basis of the original comparisons.

Much of the recent research on reading from a cognitive perspective is based on comparisons between good and poor readers. For example, in a recent article by Underwood & Zola (1986) good and poor readers were compared on letter recognition span. In this study no differences were found and no particular instructional inferences were made. In other studies this has not been the case: McGee (1982) compared good and poor fifth grade readers and poor third grade readers, finding differences in

recall of text structure ordered from good to poor fifth graders to third graders. McGee concluded that young readers "benefit from following the top-level structure of text to guide reading and remembering passage information..." Even though a disclaimer below this quote suggests the need for more research on efectiveness of instruction, there is a clear message that the observed differences are caused by what good readers do, and that poor readers will be helped by some strategy based on the good readers' processes. While the study is itself limited because of the text reading level (third grade), it is part of a chain of research related to automaticity (Laberge & Samuels, 1974) which is itself based in part on these same good-poor reader differences. There is simply no basis for assuming that the poor readers can be made to perform like the good readers or that their processing will become automatic, or if automatic in the same way that the good readers' process is automatic. Another such study is due to Sannomiya (1984) in which poor third grade comprehenders were compared with good sixth grade comprehenders on text comprehension under auditory or visual conditions. In this study both age and ability are confounded. Again, there is no evidence that the poor comprehenders can be made to look like good sixth graders, or that different modes of presentation will change reading performance in this direction.

Intact groups are also used as the basis for inferring developmental change. For example, Baldwin & Coady (1978) compared fifth graders and college students on their use of punctuation as clues to meaning in isolated sentences. They found differences between the groups and inferred developmental differ-

ences in use of punctuation as clues to meaning. It is common to see studies that mix age and reading ability. Juel (1983) compared grade two, grade five, and upper division undergraduates; good and poor readers were identified for the elementary groups. In this study the word adult was used interchangeably with the college sample, the implication being that these readers are a norm for adult performance. This assumption is most definitely wrong, and an assumption that the elementary students are likely to or can become like these adults is unwarranted. Juel nevertheless suggests that presenting children with practice words with similar letter combinations would help to develop versatility in decoding. That may be true, but the comparisons made in her study do not support such conclusions. There are many other child-adult comparisons in recent literature in which the adults are high ability college students (McGee, 1982; Schwartz, 1980; Taylor, 1980), or secondary students (Fletcher, Satz, & Scholes, 1981).

The use of intact groups has been repeatedly criticized in the educational research methodology literature from Campbell & Stanley (1963) onward with respect to the inference of causality for observed characteristics affecting behaviors. In the case of good versus poor readers, the inference is that what good readers do, poor readers can do, and that instruction directly oriented toward the discrepancy will remediate deficiencies in the poor readers. The good readers are the experts, and the poor readers the novices. The critical assumption is that the good readers were themselves in the poor readers' state at some point. Often,

since the two groups are age matched, this is not true. The good readers were never like the poor readers. Consequently, the inference that the observed differences in condition can lead to training is by itself without basis. Similarly, developmental studies are susceptible to the same difficulty, especially when they involve elementary, secondary, and college populations. In the Baldwin & Coady (1978) study a comparison between fifth grade and college students is meaningless, because differences may be due to selection: if one were able to select the fifth graders who will eventually go to college, would we still see the differences in use of punctuation clues? Even if we did find the differences, how comfortable would we be in ignoring any other differences that remain between the prospective college-bound fifth graders and the college students. Any variables upon which the two groups differ become possible alternative causal variables, and training in the absense of experimental demonstration is merely guesswork. A similar problem exists for comparisons with secondary students when dropout rate becomes appreciable (after grade 10), or when students begin self-selecting into courses (grade 9). Differential maturation and history are other threats from the Campbell & Stanley list which are relevant. Finally, regression threats due to selection of extremes are not only omnipresent in good-poor comparisons, their effects should always be estimated statistically just to provide a comparison with the observed differences.

Comparing different groups is not by itself a useless activity, but progressing directly to training is premature at best. Differences between good and poor readers, or between

developmentally different groups of readers, is useful for supplying clues or hints for more careful investigation. The tendency to assume a causal shortcut, permitting the ignoring of experimentation, is unfortunate; while the technique may prove correct in a few instances, our experience in educational research with intact groups is lengthy enough to predict many erroneous conclusions and wasted resources if the method is allowed to predominate.

Similarly, research on developmental differences has largely opted for cross-sectional designs, not wishing to do the hard research implicit in true longitudinal study of development. In the good reader-poor reader research this is particularly telling, for we have little data on long term development of either group from a cognitive, information processing theoretic perspective. This is the causal ordering problem that Rowan et al (1983) pointed out; cross-sectional designs that substitute for longitudinal designs almost always have this difficulty.

Think-aloud protocol. This technique has been used in the study of expert and novice organization of knowledge and was eloquently and favorably defended by Ericsson & Simon (1980) as a valid means to record information that humans are attending to in short-term memory. It was attacked by Phillips (1983) as an inappropriate technique to infer human's cognitive structure of subject matter. The core of Phillips' argument is that the external organization imposed in the learning required for a task may require a person to reproduce it verbally, but there is no

evidence that that organization resides internally in the person's mind. Similarly, the content and organization of a question, with perhaps the exception of free response, imposes an organization on the subject's response that does not necessarily mirror the internal representation of the response. The use of think aloud method, while it occurs in a variety of research contexts, is a major technique in naturalistic or ethnographic research. A recent study by Nicholson (1984) in which 3600 minutes of inerviewing with junior high students was conducted is an example in point. This study will be examined in more detail below, but interview techniques in the comparison of experts and novices are likely to suffer from many difficulties. In reading it is particularly problematical because the researchers usually share the same culture (reading, education, etc.) as both the experts and the novices. This is usually a drawback for ethnographers, who are attempting to view the culture with fresh eyes. in Nicholson's work the experts were teachers, and the similarity between researcher and expert was far greater than between researcher and novices (teenagers). The commonness of a shared language of educationese is quite troublesome for a researher in such conditions and the trustworthiness of such interviewing must be questioned; it is not that interviewing cannot be done well, it is that great care must be taken to support the evidence presented in such a context.

Retrospective studies. In research on creativity the comparison between creative and noncreative individuals has led to the formulation of programs to teach creativity (Van Tassel-Baska,

1986). Also, researchers on creativity have employed retrospective methods to examine prior differences between more and less creative individuals and then to propose changes in education which are expected to engender the same effects in young students as were observed in the creative adults. Segal, Busse & Mansfield (1981) compared retrospectively two groups of biologists, highly cited and nonhighly cited, using self-report survey technique. They found post-doctoral productivity to be related to pre-doctoral productivity and high school science interest.

As noted by example aove this research technique is used to infer causality by selecting groups currently differing and discovering differences in their past on putative causal variables, which are then inferred to have caused the present differences. This technique is apparently not used very much in reading research, for a search over the last ten years found only one study, by Castagna (1982) in which a historical examination of influential persons in western history was made using biographies and autobiographies. The implication is that decisive reading changed these people and that some was purposeful, some was not. Of course, historical analyses must use such methods; it is only if an implication for current practice is made that the analysis becomes suspect.

Naturalistic inquiry. This body of techniques, attempting to become a method in educational research ,in Kaplan's (1964) sense, draws upon ethnographic research from cultural anthropology, but then leaves it in a philosophical sense. Recent

apologies by Harste (undated) and Weaver (1985) liken the use of naturalisatic inquiry to a paradigm shift, citing Kuhn's now dated and largely refuted work (1963). While this debate more properly belongs in a different critical paper, the use of the techniques in the expert-novice studies requires a small aside. The appeal to a paradigm shift has been misunderstood and mislaid to boot. The shift occurred in psychology in the late 1960's and is often tied to Neisser's (1967) resurrection of internal mental representational constructs, the shift being away from behaviorism. This paradigm shift has flowed into educational research rapidly and convincingly, predating the widespread interest in ethnographic techniques by a decade. The latter interest, it is presumed, was an outgrowth of the real paradigm shift. Paradigm shifts occur in disciplines when the prevailing theories are overturned by new, revolutionary ones, that nevertheless account for the facts and relationships previously learned. In paradigm shifts the old is not discarded, it is reinterpreted. There is no such change occurring in reading, notwithstanding the wishful thinking of Harste (undated). The mistake is in confusing a change in methodology with a change in the discipline. Methodologies cannot and never will drive disciplines to the extent that the naturalistic inquirers maintain that they do; recent arguments by Kuhn (1976) himself have backpedalled on the theory-ladenness argument of data. Cooke & Campbell (1979) attack the emphasis by philosophers of science on the preeminence of theory, relegating facts to an unwarranted secondary status. That is, facts are observed by resrearchers working from different methodological perspectives. They must

reconcile them; their methodologies become more suspect than the facts, which are interobserver confirmable. If the facts are not confirmable, then they cannot be admitted. This latter issue becomes the main problem for the naturalistic researchers, for they rely heavily on impressionistic, one-shot observation for many facts. Many researchers using this method deny intersubjective confirmability, but they abandon science for art. They are not wrong, they merely inquire in another domain.

A number of naturalistic studies in reading have been published in the last few years. The study by Nicholson (1984) is the primary study I have encountered which purported to compare experts and novices. The study actually examined the structures of teenagers' understanding of classroom material; teachers were apparently ignored, although there is an appeal to teachers as experts at athe end of the study. A small section on low achievers was also tacked on. The catchy title was misleading or there was a serious editing problem because there was no comparaison between experts and novices in this study. If there had been it would have told us nothing about how to change students' conceptions. This is a common problem in naturalistic studies. One gets whatever one happens to find in the setting. If there is nothing very interesting going on little of use will be brought out. Also, naturalistic studies are limited by what passes for actual practice, not by what is possible. It is quite possible that most of what will occur in education in the next century is being tested in laboratory schools, industrial settings, and nontraditional educational locations. The public

schools are likely to be the last places to find out about these changes, whether through experimental or naturalistic means.

Naturalistic research on expert-novice differences in reading is limited by selection, ie. the choice of locations; by history, the context of the location; by instrumentation, especially changes in the observer/interviewer; and by temporal limitations in when the study is conducted and for how long. It is not argued here that naturalistic inquiry is less appropriate than the quasi-experimental research described earlier. Neither is likely to be able to draw valid conclusions regarding curriculum change in the absense of careful experimental manipulation of variables.

Summary

This paper has sought to draw attention to methodological deficiencies inherent in expert-novice research with respect to drawing inferences about curriculum change. Much credit must be given to the reading research community for generally not leaping to conclusions from such literature, in comparison with some fields of psychology, engineering, and science education. While some reading studies seem to overreach their conclusions, far more have used the observed differences to probe experimentally hypotheses generated by the observations. This approach cannot be faulted, even if one cannot resist challenging the original premise: that good readers can tell us anything about how poor readers ought to proceed. The methodological threats to internal validity of such research ventures should make us pause to consider if good-poor or expert-novice comparisons are really of

value: history, selection, instrumentation, maturation, and regression. While no study necessarily is damned due to possible internal invalidity threats, the weight of methodological argument certainly should make us pause. Ex post facto methods, such as meta analysis, can never rectify the poor initial choice of field of exploration. If we want to see how poor readers can be made into good readers we ought to find examples, or better yet, create examples, and then work to find out what is replicable. That is good science and good research.

References

Anderson, L.M., Evertson, C.M. & Brophy, J.E. (1978). The first grade reading group study: Technical report of experimental effects and process-outcome relationships, Report No. 4071. Austin:Research and Development Center for Teacher Education, University of Texas.

Baldwin,R.S. & Coady, J.M. (1978). Psycholingusitic approaches to a theory of punctuation. Journal of Reading Behavior,10, 363-375.

Brophy, J.E. (1973). Stability of teacher effectiveness. American Educational Research Journal,10,245-252.

Campbell, D.T. & Stanley, J.C. (1963).Experimental and quasi-experimental designs for research on teaching. Chicago: Rand-McNally.

Castagna, E. (1982).Caught in the act: the decisive reading of some notable men and women and its influence on their actions and attitudes.Metuchen,NJ: The Scarecrow Press.

Chase W.G. & Simon, H. A. (1973). Perception in chess. Cognitive Psychology,4, 55-87.

Chi, M.T.H., Feltovitch, P.J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices.Cognitive Science,5,121-152.

Clark, T.A. & McCarthy, D.P. (1983). School improvement in New York City: The evolution of a project. Educational Researcher,12,17-24.

Crawford, J., Gage, N., Corno, L., Staybrook, N., Mitman, A., Schunk, D., Stallings, J., Baskin, E., Hanvey, P., Austin, D., & Newman, R. An experiment on teacher effectiveness and parent assisted instruction in the third grade. Stanford, CA: Center for Educational Research at Stanford, Stanford University.

Cook, T.D. & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago:Rand-McNally.

Ericsson, K.A. & Simon, H.A. (180). Verbal reports as data. Psychological Review,87,215-251.

Fletcher,J.M., Satz, P., & Scholes,R.J. (1981). Developmental changes in the linguistic performance correlates of reading achievement. Brain and Language,13,78-90.

Harste, J.C. (undated). Portrait of a new paradigm: Reading comprehension research. In A.Crismore (Ed.), Landscapes: A state-of-the-art assessment of reading comprehension research 1974-1984. Final report, USDE-C-300-83-0130. Bloomington,IN:Indiana University Language Education Departments.

Juel, C. (1983). The development and use of mediated word identification. Reading Research Quarterly,18, 306-327.

Kuhn, T.S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Laberge, D. & Samuels, S. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology,6, 293-323.

McGee, L.M. (1982). Awareness of text structure: Effects on children's recall of expository text. Reading Research Quarterly,17,581-590.

Medley, D.M. (1977). Teacher competency and teacher effectiveness: A review of process-product research. Washington, D.C.: Aamerican Association of Colleges of Teacher Education.

Neisser, U. (1967). Cognitive psychology. New York: Appleton-Century-Crofts.

Nicholson,T. (1984). Experts and novices: A study of reading in the high school classroom. Reading Research Quarterly,19. 426-451.

Phillips, D.C. (1983). On describing a student's cognitive structure. Educational Psychologist,18, 59-74.

Rowan, B., Bossert, S.T. & Dwyer, D.C. (1983). Research on effective schools: A cautionary note.Educational Researcher, 5,24-31.

Rosenshine, B.V. (1978). Instructional principles in direct instruction. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March, 1978.

Rupley, W.H., Wise, B.S., & Logan, J.W. (186). Research in effective teaching: An overview of its development. In J.V. Hoffman (Ed.), Effective teaching of reading: research and practice. Newark,Del: International Reading Association.

Sannomiya, M. (1984). Modality effects on text processing as a function of ability to comprehend. Perceptual and Motor Skills,58,379-382.

Schwartz, R.M. (1980). Levels of processing: The strategic demands of reading comprehension. Reading Research Quarterly,15,433-450.

Segal, S.M., Busse, T.V. & Mansfield, R.S. (1980). The relationship of scientific creativity in the biological sciences to predoctoral accomplishments and experience. American Educational Research Journal,17,491-502.

Simon, D.P. & Simon, H.A. (1978). Individual differences in solving physics problems. In R.Siegler (Ed.), Children's thinking: What develops?. Hillsdale,NJ: Lawrence Erlbaum.

Taylor, B.M. (1980).Children's memory for expository text after reading. Reading Research Quarterly,15,399-411.

Underwood, N.R. & Zola, D. (1986). The span of letter recognition of good and poor readers. Reading Research Quarterly,21,6-19.

Van Tassel-Baska, J. (1986). Effective curriculum and instructional models for talented students. Gifted child Quarterly,30, 164-169.

Weaver, C. (1985). Parallels between new paradigms in science and in reading and literary theories: An essay review. Research in the teaching of English,19,298-316.